DOCUMENT RESUME

ED 395 041 TM 025 062

AUTHOR	Moore, James D., Jr.
TITLE	Stepwise Methods Are as Bad in Discriminant Analysis as They Are Anywhere Else.
PUB DATE	27 Jan 96
NOTE	21p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, January 25-27, 1996).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Computer Software; *Discriminant Analysis; Educational Research; *Error of Measurement; *Predictor Variables; Psychological Studies; Regression (Statistics); *Research Problems; *Sampling
IDENTIFIERS	Research Replication; *Stepwise Regression

ABSTRACT

The serious problems associated with the use of stepwise methods are well documented. Various authors have leveled scathing criticisms against the use of stepwise techniques, yet it is not uncommon to find these methods continually employed in educational and psychological research. The three main problems with stepwise techniques are: (1) computer packages use the wrong degree of freedom in their computations, producing spuriously statistically significant results; (2) stepwise methods capitalize outrageously on sampling error and therefore yield nonreplicable results; and (3) they do not identify the best set of predictors. As the literature already contains several examples of the misuse of stepwise methods in the case of regression, the present paper explains the problems associated with their use in the context of discriminant function analysis. It is suggested that these methods are equally as bad in multivariate statistics as they are in a univariate context and therefore should be avoided entirely. (Contains four tables and nine references.) (Author/SLD)

Running head: STEPWISE METHODS IN DISCRIMINANT ANALYSIS

Stepwise Methods Are As Bad in Discriminant Analysis As They Are Anywhere Else

James D. Moore, Jr.
Texas A&M University

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, January 27, 1996.

Abstract

The serious problems associated with the use of stepwise methods are well documented. Various authors have leveled scathing criticisms against the use of stepwise techniques, yet it is not uncommon to find these methods continually employed in educational and psychological research. As the literature already contains several examples of the misuse of stepwise methods in the case of regression, the present paper explains the problems associated with their use in the context of discriminant function analysis. It is suggested that these methods are equally as bad in multivariate statistics as they are in a univariate context and therefore should be avoided entirely.

Stepwise Methods Are As Bad in Discriminant Analysis As They Are Anywhere Else

Huberty (1994) recently noted that, "It is quite common to find the use of 'stepwise analyses' reported in many empirically based journal articles" (p. 261). Stepwise methods are typically used by researchers either to select variables to retain for further analyses or to evaluate the relative importance of the variables in a given study. It has been demonstrated, however, that stepwise methods simply are not useful for either purpose (Thompson, 1995a). Further, several authors have offered scathing criticisms of many of the common applications of stepwise techniques (cf. Huberty, 1989; Snyder, 1991, Thompson, 1989).

The present paper explains the three major problems associated with the use of stepwise methods. Although the problems delineated here are equally as pertinent in a univariate context, as in the case of regression, the focus of the present paper is on the use of stepwise methods in discriminant function analysis. Samples of results from stepwise discriminant function analyses are included in order to help make the discussion concrete.

The first major problem associated with using stepwise methods is the fact that computer packages implementing discriminant function analysis use the **wrong degrees of freedom** in their statistical tests, thereby producing incorrect results. In fact, the degrees of freedom used in

the computer packages are systematically biased in favor of yielding spuriously statistically significant results (Thompson, 1989). Although seemingly unacknowledged by most graduate students, commonly employed computer packages do not always yield infallible results.

The second major problem encountered with the use of stepwise techniques is that they tend to **capitalize outrageously on even small amounts of sampling error**, thus yielding results that will not generalize beyond the sample (Davidson, 1988; Snyder, 1991; Thompson, 1995a). If science is truly about obtaining results that can be shown to replicate under stated conditions, then it is worth asking, "Why do researchers continue to employ techniques that inhibit, or even preclude, their chances of finding reproducable results?"

The third major problem with stepwise methods pertains to the myth about what the methods actually do. Contrary to popular belief, stepwise methods **do not identify the best predictor set of a given size**. In fact, the true best set (a) may yield considerably higher effect sizes and (b) may even include none of the variables selected by the stepwise algorithm (Thompson, 1995a)! This is elaborated upon in the final section. Sample results are presented to emphasize that an all-subsets analysis is the appropriate method for determining the best predictor set of a given size.

Wrong Degrees of Freedom

Huberty (1994) states that the most popular type of discriminant analysis currently being reported is a stepwise discriminant analysis. The widespread use is undoubtedly due to the availability of computer software to accomplish the complex calculations. All three popular computer software packages - BMDP, SAS, and SPSS - include a computer program to conduct what is called "stepwise multiple regression analysis" and a program for a "stepwise discriminant analysis" (Huberty, 1989). Unfortunately, what the majority of researchers using stepwise methods fail to recognize is that the computer packages have been programmed in error and subsequently are incorrect in the number of degrees of freedom they use in their calculations.

Thompson (1994) reminds us that degrees of freedom are like coins that we can spend to investigate what's going on within our data, i.e., what explains or predicts the variability in the dependent variable(s). Each time a predictor variable is "used" in the analysis, there is a "charge" of one degree of freedom (explained). In a stepwise discriminant analysis, or any other stepwise procedure for that matter, the computer packages are programmed to "charge" us one degree of freedom each time a new variable is included in an analysis (i.e. at each "step" in a forward selection procedure). In actuality, however, all predictor variables in our original variable set are

involved in each step in that each of them is considered for inclusion. Therefore, the correct number of degrees of freedom that should be "charged" is the same at each step and is equal to the total number of variables in the predictor set. Obviously, this "additional charge" will dramatically decrease the likelihood of obtaining statistically significant values (i.e. by decreasing the F ratio thereby increasing p-calc).

For example, let's say that a researcher is conducting a stepwise discriminant analysis for a design in which he/she is attempting to find variables that best explain the differences between three groups based on a set of ten response variables. After much thought, he/she decides that it would be optimal to "whittle down" the response variables to the three with the most explanatory power. Therefore, the analysis is run (on any of the common statistical packages) and is complete after the third step. If this researcher believes the significance values reported for each step of the analysis to be correct, then he/she is destined to make grave errors regarding the overall explanatory power of the three variables selected. Other considerations (which are addressed later in this paper) notwithstanding, the explanatory power of each of the variables is not accurately reflected in the significance values reported at their respective steps due to the computer packages inaccurate use of the degrees of freedom.

At each step in this analysis, the degrees of freedom (explained) should have been computed as ten. However, the way in which the computer packages have been programmed would have led the computations of the degrees of freedom (explained) at each of the three steps to be one, two, and three respectively. For this reason, the reported significance values for the variables are in error and are (systematically) biased in an upward direction. Without knowledge of the computer package's error, this researcher is likely to conclude that the variables in his/her analysis contain far more explanatory information than they actually do. Both Snyder (1991) and Thompson (1995a) offer detailed explanations of the ways in which the computer packages' use of the wrong degrees of freedom can impact the results one obtains.

Capitalizing on Sampling Error

A far more serious problem than the degree of freedom issue (which can be corrected for by hand) in the use of stepwise methods relates to the way in which these methods ten to capitalize outrageously on even small amounts of sampling error thereby producing results that are not replicable. A stepwise analysis is unique to other types of analyses in that in considers variables for inclusion in the analysis one at a time and in the context of previously entered variables (of course the reverse is true in a

backward selection approach). Thompson (1995a) states that stepwise analysis is a linear series of conditional choices not unlike the choices one makes in working through a maze. An early mistake in the sequence will corrupt the remaining choices. That is to say, there are likely to be cases in stepwise analyses where one variable is chosen ahead (i.e. at a prior step) of another due to an infinitesimal advantage. The question then arises as to whether or not that slight advantage constitutes a true superiority on the part of the chosen variable or an advantage simply due to random variance?

> At a given step, the determination of which single variable to enter will enter variable X1 over variable X2, X3, and X4, even if X1 is only infinitesimally superior to the other three variables. It is entirely possible that this infinitesimal advantage of variable X1 over another variable is sampling error, given that the competitive advantage of X1 is so small (Thompson, 1995a).

Further, given the nature of stepwise methods, where variables not included in the analysis on a given step are evaluated in terms of their ability to contribute unique explanatory information to those variables already included in the analysis, it is possible that otherwise worthy variables are often excluded from the analysis altogether.

In such a case, many researchers may erroneously conclude that variables not included in the analysis contain no explanatory or predictive potential. In fact, this may not be the case at all, and such a conclusion cannot be drawn from merely conducting a stepwise analysis. Variables excluded from the analysis through the stepwise algorithms may contain much potential for explaining group differences but may not contribute enough unique information to the variables included prior in the analysis. As stated above, this issue takes on a great deal of importance when one considers that a given variable may be chosen ahead of another due to sampling error alone.

Insert Table 1 About Here

To make the discussion more concrete, partial results from a stepwise discriminant function analysis are presented in Table 1. In this case, there are four repsonse variables (Y1, Y2, Y3, and Y4) from which we are trying to describe the differences between three groups. Only two functions are presented in this case to keep the discussion as simple as possible. From the standardized canonical discriminant function coefficients listed in Table 1 it is apparent that variable Y1 is receiving most of the explanatory "credit" on the first function while variable Y3 is receiving the credit likewise on function two.

Insert Table 2 About Here

To conclude at this point, however, that only variables Y1 and Y3 have explanatory potential would be premature. A glance at the structure matrix presented in Table 2 illustrates that both Y1 and Y2 have high correlations with function one and that Y3 as well as Y4 correlate highly with function two. These two tables, taken together, suggest that while both variables Y2 and Y4 may have a great deal of potential in terms of describing the differences in the three groups on function one and function two respectively, variables Y1 and Y3 are receiving the credit. This is so because variables Y1 and Y2 are likely highly correlated with one another as are variable Y3 and Y4. Due to the high degree of these correlations, variables Y2 and Y4 offered little unique explanatory information to the analysis after variables Y and Y3 had already been entered and therefore were assigned low weights.

Remember, however, that the small differences in explanatory power between Y1 and Y2 and between Y3 and Y4 could have been due to sampling error in which case these results are not likely to replicate. In fact, in future attempts at replication, it would not be unlikely to see variables Y2 and Y4 receive the credit for differentiating the groups on functions one and two respectively.

Not Selecting the Best Subset

Huberty (1994) states that most researchers who employ stepwise methods in their analyses do so primarily for two reasons: 1) to select variables to retain in an analysis, and 2) to order the variables in terms of their relative contributions to the analysis. Of course, it has been shown that stepwise methods, either in univariate or multivariate contexts, do not provide accurate results for either purpose (Snyder, 1991; Thompson 1989 & 1995a). The problems associated with using stepwise techniques in discriminant analysis for the purpose of ordering variables was discussed in the previous section. In sum, due to stepwise methods' tendency to capitalize on even small amounts of sampling error, the step at which a variable is included in an analysis may not at all reflect that variable's "true worth." The problem of using stepwise methods to select variables to retain in an analysis is the focus of the present section.

In using stepwise techniques for the purpose of selection (i.e. choosing a subset of variables from the original variable set), a researcher has failed to recognize the basic question that stepwise techniques are designed to answer. The stepwise algorithms are written so as to evaluate the relative unique contribution of variables one at a time. At no point in their computations do stepwise techniques ever ask the question, "What is the best subset

of predictors of a given size?" It is a grave error in logic, then, to conclude that one has received an answer (from the results of a stepwise analysis) to a question that he/she has not posed (Thompson, 1995a).

Insert Table 3 About Here

Table 3 presents partial results of a stepwise discriminant analysis procedure. In this example, there are ten response variables which are being used to describe the differences between a number of groups. The top portion of the table lists the variables along with their corresponding F to Enter and Wilks' Lambda values prior at step 0. Let us say that we are interested in selecting from this original set cf ten, the "best" subset of size three. Therefore, our analysis is complete after three steps - the results for which are presented in the bottom of table 3.

From these results, it appears as though the "best" subset of size three from our original set of ten consists of variables Y1, Y2, and Y3. This is where many reserachers draw erroneous conclusions. While it may be true that variables Y1, Y2, and Y3 each offer worthy information to our analysis, how can we be certain that they, in actuality, constitute the **best** subset of size three? Of course, we cannot make that conclusion since the stepwise algorithms

are not set up to evaluate subsets. Rather, the decisions made in a stepwise analysis regarding whether or not to include variables in an analysis are made in a linear sequence fashion within which each variable is evaluated independently in the context of the presence of the other variables. Thompson (1995a) offers a literal analogy to this situation:

> Suppose one was picking a basketball team consisting of five players. The stepwise selection strategy picks the best potential player first, then the second best player in the context of the characteristics of the previously selected player, and so forth. An alternative selection strategy is an all-possible-subsets approach, which asks, "which five potential players play together best as a team?" This team might conceivably contain exactly zero of the five players selected through the stepwise approach and might be able to stomp the "stepwise team."

Insert Table 4 About Here

Table 4 presents some sample results from an all-possible-subsets approach for the variables that were listed in Table 3. It so happens in this case, that the

best subset of size three turns out to be Y2, Y4, and Y5. Recall that the stepwise procedure had selected Y1, Y2, and Y3. The all-possible-subsets approach reveals that a subset consisting of these variables is not only not the best subset but that there are four better subsets. Although these results are fictitious and are for heuristic purposes only, given the nature of the stepwise selection process it is reasonable to expect different (sometimes dramatically different) results when selecting via an all-subsets-approach. Huberty's (1994) book is accompanied by a computer diskette which contains all-subsets-approach programs by both Morris and McHenry. Also, the RSQR procedure in SAS can be used to analyze all possible subset combinations.

One final problem with using stepwise methods for selecting variables in a discriminant analysis context has to do with the criterion on which variables are chosen. The Wilks' Lambda statistic is what the computer packages base their decisions on in deciding whether to add variables in a given analysis. This is to say, as the variables being considered are evaluated, the computer is programmed to select the one variable (at a given step) which offers the greatest contribution to the Wilks' Lambda value (i.e which one reduces it the most). Huberty (1987) reminds us that while this selection criterion may be appropriate in a descriptive discriminant analysis case (where the focus is

on explaining group differences), is seems inappropriate in a predictive discriminant analysis where the focus should be on correctly assigning subjects to groups. Although some researchers may argue that separating groups is tantamount to being able to accurately assign subjects, Thompson (1995b), offers a detailed explanation of why this not necessarily so. In fact, it is demonstrated in that article that the number of correct classifications may actually **decrease** in a predictive discriminant analysis when Wilks' Lambda is used as the criterion for determining additional variables to include in the analysis. The criterion of interest in a predictive discriminant analysis should be the "hit rates" one obtains, not simply a decrease in the Wilks' Lambda statistic.

Conclusion

A great deal has been written about the misconceptions and misuse of stepwise methods. At this point, however, it appears that they are continually being employed in psychological and behavioral research. The three main problems with stepwise techniques are as follows: 1) computer packages use the wrong degrees of freedom in their computations thereby producing spuriously statistically significant results, 2) stepwise methods capitalize outrageously on sampling error and therefore yield non-replicable results, and 3) they do not identify

the best subset of predictors, contrary to what many may beleive. The primary intent of the present paper, therefore, has been to further persuade researchers against using stepwise methods altogether in lieu of more appropriate alternatives. It should be clear at this point, that stepwise methods are equally as bad in discriminant analysis as they are anywhere else.

References

Davidson, B. M. (1988, November). The case against using stepwise research methods. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville. (ERIC Document Reproduction Service No. ED 303 507)

Huberty, C. J. (1989). Problems with stepwise methods: Better alternatives. In B. Thompson (Ed.), Advances in social science methodology (Vol. 1, pp. 43-70). Greenwich, CT: JAI Press.

Huberty, C. J. (1994). Applied discriminant analysis. New York: Wiley and Sons.

Huberty, C. J., Wisenbaker, J. M., & Smith, J. C. (1987). Assessing predictive accuracy in discriminant analysis. Multivariate Behavioral Reserach, 22, 307-329.

Snyder, P. (1991). Three reasons why stepwise regression methods should not be used by researchers. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 99-105). Greenwich, CT: JAI Press.

Thompson, B. (1989). Why won't stepwise methods die? Measurement and Evaluation in Counseling and Development, 21(4), 146-148.

Thompson, B. (1994, April). Common methodology mistakes in dissertations, revisited. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Thompson, B. (1995a). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. Educational and Psychological Measurement, 55(4), 525-534.

Thompson, B. (1995b). Review of Applied discriminant analysis by C.J. Huberty. Educational and Psychological Measurement, 55, 340-350.

Table 1. Standardized Canonical Discriminant Function Coefficients

Variable	Function 1	Function 2
Y1	.70835	.10132
Y2	.21364	.12783
Y3	.08214	.71863
Y4	.11267	.24632

Table 2. Structure Matrix

Variable	Function 1	Function 2
Y1	.92435	-.08723
Y2	.90245	.07256
Y3	.12865	.90765
Y4	-.09873	.88546

Table 3. Sample of Stepwise Selection Procedure

-------Variables not in the analysis after step 0-------

Variable	F to Enter	Wilks' Lambda
Y1	128.6543	.2573240
Y2	110.8654	.3126234
Y3	90.8762	.4076238
Y4	75.9282	.4592876
Y5	68.8272	.5198287
Y6	54.8376	.6582028
Y7	45.9828	.8097132
Y8	16.2882	.9245462
Y9	10.5626	.9562811
Y10	5.9342	.9842561

-----------Variables in the Analysis after Step 3-----------

Variable	F to Remove	Wilks' Lambda
Y1	35.1185	.2159544
Y2	30.4556	.2070129
Y3	26.7258	.1841180

Table 4. All-Possible-Subsets Results (of size three)

Variable Subset	Wilks' Lambda
Y2, Y4, Y5	.1186752
Y2, Y4, Y1	.1562869
Y1, Y3, Y5	.2087266
Y2, Y3, Y5	.2172653
Y1, Y2, Y3	.2462983
Y1, Y2, Y4	.2783936
Y1, Y2, Y5	.3274522
Y3, Y4, Y5	.3689278
Y1, Y4, Y5	.3965283
Y2, Y5, Y6	.4293752
etc.	